

PROPERTY OF
PRINCETON UNIVERSITY LIBRARY
RECEIVED JAN 6 1938

The PSYCHOLOGICAL RECORD

DECEMBER, 1938
Vol. II No. 17

USE AND ABUSE OF MENTAL TESTS IN CLINICAL
DIAGNOSIS

GRACE H. KENT



THE PRINCIPIA PRESS, INC.
BLOOMINGTON, INDIANA

Price of this number, 25 cents

EDITOR

J. R. KANTOR, *Indiana University*

BUSINESS MANAGER

C. M. LOUITT, *Indiana University*

DEPARTMENTAL EDITORS:

ABNORMAL	EDMUND S. CONKLIN, <i>Indiana University</i>
CHILD	HELEN KOCH, <i>University of Chicago</i>
CLINICAL	B. M. CASTNER, <i>Yale University</i>
COMPARATIVE	E. A. CULLER, <i>University of Rochester</i>
EDUCATIONAL	J. G. PEATMAN, <i>City College of New York</i>
EXPERIMENTAL	B. F. SKINNER, <i>University of Minnesota</i>
MEDICAL	NORMAN CAMERON, M. D., <i>Cornell Medical College</i>
PHYSIOLOGICAL	C. F. SCOFIELD, <i>University of Buffalo</i>
PSYCHOMETRICS	J. P. GUILFORD, <i>University of Nebraska</i>
SOCIAL	NORMAN C. MEIER, <i>University of Iowa</i>
ASSISTANT EDITOR	J. W. CARTER, JR., <i>Indiana University</i>



The Principia Press, Inc., has undertaken the publication of this co-operative journal to afford authors the opportunity of immediate publication at the least possible expense. The present low cost of publication, and possible future reductions, depend entirely upon the number of subscriptions. The subscription price has been made low to induce individuals to subscribe. Under the Articles of Incorporation of the Principia Press no profit can be made on any of its publications. Therefore an increase in the number of subscribers will be reflected in reduced costs to authors and in increase in the number of pages published annually.

EDITORIAL BOARD. The above named board of associate editors have agreed to serve in an active capacity for a period of three years, and all manuscripts will be submitted to them.

MANUSCRIPTS may be sent to the appropriate associate editor or to Dr. J. R. Kantor. Longer papers (30 or more typewritten pages) in the above mentioned fields are especially desired.

COSTS. The cost of publication will be \$2.00 per page of approximately 470 words. The charge for line cut illustrations will be \$2.00 regardless of size. Special charges will have to be made for complex tabular matter and for half-tone or colored illustrations.

REPRINTS. One hundred copies of each paper will be supplied gratis. Additional copies may be purchased in any quantity at a rate which will be set when the author is informed of the cost of publication.

ROYALTIES. Fifty per cent of the net income from the sale of individual copies, or from copies sold as part of back volumes, will be credited as royalties to the author's account. Royalties cannot be paid on income from subscriptions current in the year of publication.

SUBSCRIPTIONS. The subscription price is \$4.00 per year for an annual volume of approximately 500 pages. Individual papers will be sold at an advertised price, which will depend upon the size of the article. Foreign subscriptions will be \$4.50.

CORRESPONDENCE concerning editorial matters should be addressed to Dr. J. R. Kantor, Indiana University. Business correspondence should be addressed to Dr. C. M. Louttit, Principia Press, Inc., Bloomington, Ind.

USE AND ABUSE OF MENTAL TESTS IN CLINICAL DIAGNOSIS*

BY GRACE H. KENT

Danvers State Hospital, Hathorne, Massachusetts

The boom which applied psychology enjoyed during the years immediately following the World War has been unfavorable both to the development of carefully-thought-out testing methods and to the proper use of such methods as have been developed. It is accepted by the general public as almost axiomatic that a child's IQ can be determined with reasonable accuracy, and it is expected that the annual report of any institution for children shall contain full data concerning the intelligence of the children. The Stanford-Binet scale is recognized by some state legislatures as constituting a birth registry for the determination of mental ages. Children have been certified for commitment to institutions for feeble-minded solely or primarily on the strength of low rating by tests.

No one who has never witnessed the presentation of the Binet scale can fully appreciate to what extent the validity of the results depends upon the mood of the subject; nor can anyone who has not presented the test in person guess how many of the subject's failures may be due to the examiner's headache or fatigue. The physician who makes use of the test results for aid in diagnosis and recommendations is not usually familiar with these sources of error. It is rather exceptional for a physician to give much attention to the technique of psychometric examination, because he looks upon it as a piece of clerical work with which he need not concern himself. As a rule, the untrained psychometrist who hands over the numerical results without comment is at a premium as compared with the more experienced examiner who dares to challenge the validity of the findings; and the psychometrist who can complete three or four examinations in two hours is rated higher by the physician than the more careful worker who refuses to be hurried. Thus there is very little to encourage the psychometrist either to do the work as well as possible or to present an honest report of the examination.

There is an increasing demand for psychometric findings which can be used in statistical compilation; so the physician himself is under some pressure to obtain the desired information from the psychometrist. There seems also to be an increasing tendency to use the findings unconditionally

* Presented in substance for discussion at Round Table, Columbus, September 9, 1938. Recommended for publication by Dr. C. M. Louttit, November 25, 1938.

in the disposition of a given case, as is illustrated by the following incident:

A public school officer had established a ruling that any child whose rating by Stanford-Binet indicated a three-year mental retardation should be placed in special class, whatever his age or school achievement. A subject who came under this ruling was a boy of 14 years who was doing fair work in sixth grade. His Binet rating was only 9-4, but it was reported that he responded carelessly and without interest. The teachers did not consider him a suitable case for special class, so they referred him for a more thorough examination. It was found that he did better work in written tests than in orally presented tests, which indicated both that the Binet rating was too low and also that the boy had passed the developmental level for which the special class is intended. In a series of nine tests, each of which yielded a rating of at least 10 years, he achieved a median of 12 years. This reduced his mental retardation to two years, and the results of the irregular examination were accepted to the extent of permitting the boy to remain one more year in the regular grades; but only with the understanding that he should be placed in special class the following year if he should retain his 12-year rating after reaching the age of 15.

Presumably no psychologist would wish to see test findings used so mechanically as this school officer is using them, but we cannot disclaim the responsibility. We have come forward with an offer to furnish intelligence ratings at wholesale rates. We have led the public to overestimate the degree of accuracy with which mental capacity can be determined. We have claimed too much for our tests, and have been taken at our word. The situation is essentially one of our own making.

An important step toward making the mental test a safe and useful instrument in clinical examination is to break down the undue confidence which the public places in the findings. To this end certain ideals which will not be realized in the near future are offered for the consideration of the younger students. Any progress we can make toward them under present conditions will have to be unsteady and inconsistent; and yet it seems worth while to keep them in mind as goals to be approached.

I. REPUDIATION OF THE CLAIM THAT WE CAN MEASURE INTELLIGENCE

It is desirable to discontinue the use of the term "intelligence" as applied to anything that can be measured.

Certain particular aptitudes can be measured crudely by tests in current use, and the complex which we call intelligence doubtless includes some measurable apti-

tudes along with other aptitudes which—if measurable—have not yet been measured. But why should the term "intelligence" be restricted to what we can measure? We need the word occasionally when referring to a person who has not been tested. In clinic conferences it is a useful term in describing the attitude of a child's mother and in considering what sort of cooperation may be expected from her; and when we speak of this mother as being an intelligent person we have in mind something quite other than the score she might possibly achieve in a test. Furthermore, if we appropriate the word for describing what is measured by tests, the convenience of the one-word term tempts us to talk about it rather more freely than our knowledge justifies. We speak of one test as an intelligence test, and refer to some other test a little disparagingly as an information test, an aptitude test or an achievement test,—just as if we could isolate intelligence from information and from special aptitudes. Such careless use of language may lead to loose thinking. A term requiring a little more effort—such as "measurable mental capacity" or "the ability measured by this test" would be conducive to greater care in drawing conclusions concerning what is measured by tests.

II. RECOGNITION THAT OUR ABILITY TO MEASURE WHAT CAN BE MEASURED IS ONLY RELATIVE

Our norms hold only for the group, not for every individual included in the group and not necessarily for any particular individual. We should make it known that we can no more guarantee to ascertain the measurable mental capacity of each and every child referred to the clinic than the actuary can name the exact year in which a given man will die.

It is true that the test contains many items, whereas actuarial tables are based upon one single item. However, the parallel holds in that we depend upon numbers for whatever validity we claim for our norms. In general, our confidence in the norm varies directly with the number of cases which it includes; and when we offer a test that is standardized upon a small number of subjects, we do it apologetically. A wide distribution of individual scores at any given age or mental level is assumed as a matter of course, and we aim always to have a large enough number of cases to yield a representative average for each mental level covered by the test. Up to this point our method of developing norms has much in common with that of the insurance company in establishing vital statistics, but here we part company with the statisticians. After our norm has been established, we proceed to the assumption that it holds for the individual as well as for the group. Of course we may feel reasonably confident that it holds for the majority of cases, but there

is always a chance that any particular case under consideration may be one of the exceptional cases for which the norm has no validity at all. It is not safe to stake anything important upon the applicability of the norm to a given case unless the findings are supported by additional evidence.

III. CRITERIA FOR INDIVIDUAL ACHIEVEMENT RATHER THAN FOR GROUP ACHIEVEMENT

The criterion for a test recommended for use in clinical diagnosis should be *the degree of uniformity which it yields when applied to a group of children who are known to be well-matched in school achievement.*

A suitable group of subjects for the test-group can be found in a finely graded urban or suburban school, by selecting from a given middle section all those children whose ages are within six months of the average age for that section. It might be well further to select a home-owning community, and to limit the group to those children who have been under identical school instruction for at least two years. The test-group need not be a large one, and therefore it would be permissible to use any desired basis of selection for the town, the school and the grade; but there should be no individual selection of children within the group.

Until we have a test battery that will yield passably uniform results under conditions highly favorable to uniformity, it is obviously unfair to expect every child examined in the clinic to conform to a statistically-established standard in which no account is taken of individual differences. Even if we had a test which would meet this requirement for normal children, it would not necessarily be applicable to every clinic child.

Any particular test unit may be expected to show wide individual variation; but it should be our aim to assemble a combination of tests so varied in nature as to give any subject—whatever his individual strength and weakness—a fair opportunity to show what he can do. This was Binet's ideal, and it has doubtless been the goal of everyone who has undertaken to establish a mental measurement scale. Unfortunately, a composite scale does not permit us to judge with what success the goal has been approached. It is only when we have independent norms for each test unit that we can try out different combinations for the purpose of ascertaining whether we are measuring a wide range of aptitudes or merely measuring the same aptitude in different ways.

In assembling material for a well-balanced battery, the first step is to find test units of low intercorrelation. Much has been written on correlation among tests, and a vast amount of effort has been expended in proving that some

new test shows high correlation with some well-established test; but too little attention has been paid to our need of tests which measure something distinctive. A good problem for mathematically-minded students is to arrange selected test units in pairs showing the minimum of correlation within each pair. (A pair showing negative correlation is a little too much to expect.) With a generous supply of units thus paired, it might be possible to make up a battery which would yield a wide range of individual ratings for almost any given subject. The wider the scatter among these ratings, for any particular subject, the more trustworthy the battery as a whole. It is the median rating by the battery that should show approximate uniformity for the selected group.

This is a project for the future. In the meantime, we owe it to the children whom we examine in clinics to be very cautious about offering test findings that may be used in individual diagnosis.

IV. DRASTIC MODIFICATION OF THE CONCEPT "IQ" AND OUR METHOD OF DERIVING IT

When we report a child of 11 years as having an IQ of 83, it should mean exactly what it is popularly believed to mean: that his achievement in the test is 83 per cent of the average achievement—empirically determined for this particular test—among children 11 years of age.

Among psychologists it is widely conceded: that the current method of deriving the IQ is crude and arbitrary as compared with the method of ascertaining the "mental age"; that mental development proceeds more rapidly in early childhood than in adolescence; and that the conversion of the "mental age" into the IQ discriminates against the older child as compared with the younger one. But the majority of persons to whom the psychometric reports are submitted, being unfamiliar with the technique, naturally assume that the IQ is essentially as sound as the "mental age." They prefer the IQ because of its convenience, and are usually incredulous when informed that according to our system of deriving this figure the adolescent of fifteen is held responsible for developing as rapidly as the child of five.

When a child's "mental age" happens to be the same as his life age, it does not matter how the findings are expressed or in what way the IQ is derived. This child's IQ, derived by whatever method, will of necessity be exactly 100. But the wider the discrepancy between the age and the test rating, the more inaccurate is the current method of deriving the IQ. This holds for deviations in either direction, but in the clinic it is the retarded child with whom and for whom we are especially concerned.

If a test were so standardized as to yield adequate norms for each age covered by the test, so that the score might be converted directly into a percentage of average achievement for the subject's own age, there would be no objection to this rating beyond such objection as applies also to the "mental age" and all other statistical ratings. Whether anything less than a true percentage of achievement at the age level of the subject should be accepted as an equivalent of such rating is an open question. (Being frankly afraid of any short-cut method of obtaining a percentage rating via the "mental age," and especially because of having learned many years ago to dispense with the IQ entirely for personal use, the writer has not made sufficient use of the Hilden tables (1) to be prepared either to endorse or to criticize this method of obtaining the IQ. The criticism here expressed refers to the original Stanford-Binet (2) method, which has been in use since 1916 and which we are still required to use in reporting cases of certain types).

It is freely conceded that the one-figure rating is very convenient, for reports on children in the developmental period; but it does not follow that its convenience is of such importance as to justify the practice of reporting findings that are grossly misleading. The physician and probation officer can learn to think in terms of the "mental age," just as they had to do before the IQ came into general use. Knowing a child's actual age, they can see at a glance whether his test achievement is above or below that age. The fact of retardation or advancement is more significant than the degree. Surely it is better to report such findings as can be offered honestly than to report a figure which has the appearance of being precise but which in many cases is actually false.

V. ABSOLUTE ABANDONMENT OF THE "IQ" OR ITS EQUIVALENT FOR ADULT SUBJECTS WHO ARE RATED BY NORMS DERIVED FROM CHILDREN

This is recommended not only because we cannot agree what age should be accepted as the normal adult level, but also because the "mental age" rating is quite sufficient as it stands.

It is for children in the period of rapid mental development that the IQ rating possesses any considerable advantage over the "mental age" rating. The IQ is undeniably convenient as a means of reducing the child's achievement to one figure, but at the adult level there is not even this excuse for the inaccuracy which the IQ involves. For the adult, the "mental age" rating is already in one figure; for the adult, the absolute level of mental development as measured by the test is the most significant figure to be reported. The IQ is therefore as unnecessary as it is questionable.

This is true for the adolescent as well as for the mature person, and it is a vastly more important consideration for our numerous adolescent subjects than for the occasional adult subject. The age of 13 years is recommended as the limit beyond which there is nothing to be gained by converting the "mental age" into the IQ. This must not be understood as implying that 13 years is the age limit of measurable mental development. We need not set any age limit at all to mental development, nor need we set the limit of measurable mental development anywhere nearly so low as 13 years. This age, however, does seem to mark the point beyond which mental measurement becomes increasingly difficult and accordingly more uncertain. Children begin in early adolescence to show more individuality in their interests and wider divergences in their aptitudes; so it naturally becomes increasingly difficult to devise test methods which approach universal applicability. The writer has standardized tests which show a clear year-to-year graduation up to 13 years, but not one which differentiates sharply between 13 and 14 years. For the child over 13 years of age, as for the adult, the absolute mental level reached by the subject is of such significance that there is no need of expressing it in any form other than the "mental age" rating.

There is of course no objection to the percentage rating for a subject of any age, if it be based upon norms obtained from unselected persons of approximately the subject's own age. It is questionable, however, whether this ideal plan of test standardization would be worth while beyond the age of 13; and it is doubtful whether it would be even possible beyond the age of compulsory school attendance.

VI. RECOGNITION OF THE "DOUBTFUL" REPORT

It seems almost beyond belief that a physician should be expected to affix his signature to a psychometric report which has been repudiated by the examiner, but this requirement is an unavoidable part of the system under which we are working. Not often is any use made of the invalid findings. The report is primarily for record, and in most cases it is merely buried in the files. But although the practical effect of placing an invalid report on permanent record may be negligible in the vast majority of cases, it is still a source of potential danger that a child of low rating may be denied some educational opportunity to which he is fairly entitled, such as being admitted to a trade school. Also, it is nothing less than destructive to the self-respect of the psychometrist who is required to report the results of an unsatisfactory examination. Both for the protection of the child and for our own standards of intellectual integrity, we should insist upon the privilege of withholding the numerical findings

whenever the examiner is satisfied that they do not fairly represent the actual ability of the subject.

The laboratory technician is not expected to report a Wassermann reaction as positive merely on the ground that it just misses being unmistakably negative, but is expected rather to repeat the observation. But it is much easier to obtain a new specimen of blood for a repeated Wassermann test than to obtain a second interview with a clinic patient. Also, there is no guarantee that a child will give better co-operation at the second interview than at the first one, so the practice of repeating the examination would not invariably furnish a solution. We cannot promise in a given case that the report will have sufficient validity to justify its being placed on permanent record. This, however, does not offer any reason for the acceptance of an invalid report.

The safest way to avoid having the records so loaded with incorrect data as to make them statistically worthless is to have a recognized place for records in which the "mental age" and the IQ are reported as "undetermined."

VII. INTERPRETATION OF PSYCHOMETRIC FINDINGS BY ONE WHO KNOWS THE TECHNIQUE OF EXAMINATION

It seems obvious that the responsibility for interpreting the results should not be placed upon a person who is wholly unfamiliar with the technique, but this procedure is by no means uncommon. Many current abuses are traceable to the plan of letting a slightly-trained psychometrist report directly to a technically-uninformed physician or probation officer.

Inasmuch as no one else can interpret the findings quite so well as the examiner, it is desirable that the examination be made by a person who is qualified to interpret the results. This, however, is not always possible. If the routine work of the clinic must be done by a technical assistant, provision should be made for having the results of each examination analyzed by an experienced examiner before the report is submitted to the person responsible for the disposition of the case.

VIII. RECOGNITION OF SUBJECTIVE EVALUATION OF TEST RESULTS

There is a tendency to exaggerate the objectivity of the psychometric test as individually presented in the clinic. It is not quite so objective as we like to consider it, and not nearly so objective as the non-psychological public assumes. We can control—to some limited extent—the conditions under which the test is presented; but not the conditions under which it is received. While we are struggling to make the presentation as objective as possible, the subjective factor creeps in by the back door.

A group examination is more nearly objective than an individual examination, but we do not for that reason consider the group examination the more trustworthy. The Binet examination could be presented more uniformly and objectively by a phonograph than by a living examiner, but the subject would not be expected to react normally to so unnatural a situation. We do not go quite to the limit of attainable objectivity, but we do go so far in that direction as to leave behind something of value in the examination. The test is a very useful instrument, but it is not an instrument of precision. While our attention is focused upon such exact determinations as can be wrung from the test, we sacrifice something of the opportunity it offers for studying the reactions of the subject for what we can learn from them.

It is important to keep clearly in mind the distinction between the objective and the subjective factors, so that we may not deceive ourselves concerning the measure of objectivity which is attained.

One method of teaching young psychometrists to observe and to write descriptive reports is to require them occasionally to present without the stop watch some well-known performance test which is scored only by speed. The way in which a subject reacts to the increasing difficulty of a series may be incomparably more significant than the score, but it may escape the notice of the examiner who is interested in obtaining a score. The deadly monotony of presenting the same test day after day may be varied by the supplementary use—when time permits—of an unstandardized test which can be evaluated only by observation; and the psychometrist may thus be kept alive to the possibilities of using tests intelligently instead of mechanically. In the reports, however, the inexperienced examiner must be taught to differentiate sharply between observation and opinion.

The most significant clinical observation ever made by the writer came about quite by accident, when a new form-board series was the center of interest in the department. The patient—a case of mental deafness which was probably congenital—had been diagnosed as an idiot on the ground that he had no understanding of language. Here was an opportunity to try out the new test on a boy who did not recognize his own name. But his formboard performance was not that of an idiot, and the interest of the two examiners was quickly shifted from the test to the subject. Further observations were made, on the strength of which he was sent to a school for the deaf. Six months later it was reported that he was making fair progress in learning to read and that he had written an intelligible letter to his mother. That boy might have gone through life without a language, except for the accident of our having at hand a

new toy that we could not resist playing with. Not since the test was standardized has it yielded results of far-reaching importance. We have fallen into the habit of depending upon the norms to save us the trouble of making observations.

The professional clinical psychologist should have tests which constitute his own personal tool, the results of which may be evaluated by his own subjective norms. They may or may not be tests of his own devising, but it is essential that they be to his own individual liking. There are those who like the Stanford-Binet vocabulary test well enough to use it habitually for opening any examination and for establishing a friendly contact with the subject. If they can thus take over as a personal instrument a test which is part of a prescribed examination, it is so much to the good. But the writer prefers to open an examination with some task which is not standardized, which need not be presented according to any fixed rules, and which may be interrupted if the subject's interest in it fails to develop. It may require several trials to find something that will stimulate the interest of the adolescent boy who deeply resents having been ordered by the court to take a "nut" examination. In order to overcome the hostility of a difficult subject, it is convenient to have within reach a collection of highly varied tasks, each of which possesses some inherent challenge. These unstandardized tests usually serve as the introduction to a standard examination rather than as a substitute for it; but they frequently constitute the most vital part of the examination and yield the most helpful part of the report.

It may take another generation of servile subjection to rules and fetishistic confidence in statistically established criteria before clinical workers in the large will be encouraged or allowed the time to find out what tests are good for; but, with the ever-increasing supply of test material to draw from, there is a chance that more and more of us may discover what magnificent possibilities the mental test holds for individual study.

REFERENCES.

1. Hilden, A. H., *Table of Percent of Average Development Based on Mental Growth Units*. Minneapolis, Educational Test Bureau, 1936.
2. Terman, L. M., *The Measurement of Intelligence*. Boston, Houghton Mifflin, 1916.

